COLECCIÓN **KUAA**

# ¿APRENDIZAJE AUTOMÁGICO?

## UN VIAJE AL CORAZÓN DE LA INTELIGENCIA ARTIFICIAL CONTEMPORÁNEA

ENZO FERRANTE (DIRECTOR)
LAURA ALONSO ALEMANY
DIEGO FERNANDEZ SLEZAK
LUCIANA FERRER
DIEGO MILONE
GEORGINA STEGMAYER

PRÓLOGO: DIEGO GOLOMBEK



**VERA** editorial cartonera

# ¿APRENDIZAJE AUTOMÁGICO?



# ¿APRENDIZAJE AUTOMÁGICO?

## UN VIAJE AL CORAZÓN DE LA INTELIGENCIA ARTIFICIAL CONTEMPORÁNEA

COLECCIÓN **KUAA**

ENZO FERRANTE (DIRECTOR)
LAURA ALONSO ALEMANY
DIEGO FERNANDEZ SLEZAK
LUCIANA FERRER
DIEGO MILONE
GEORGINA STEGMAYER

PRÓLOGO: DIEGO GOLOMBEK

**VERA** editorial cartonera

Pedimos perdón
Corriendo, enmascarando el fin
Por eso te busqué, por eso diseñé
La máquina de ser feliz

Plateada y lunar
Remotamente digital
No tiene que hacer bien, no tiene que hacer mal
Es inocencia artificial

CHARLY GARCÍA, «La máquina de ser feliz»

# PRÓLOGO

DIEGO GOLOMBEK

Mi inteligencia intrapersonal
nunca ha sido la mejor de este pueblo

INTOXICADOS

Posiblemente la gran pregunta de todas las ciencias es, sencillamente, qué es esto que somos. ¿Un grupo de células, instrucciones genéticas, el azar y la necesidad de estar vivos? ¿Una coctelera con 59 elementos de la tabla periódica, bien mezclados y en su justa medida? A ver: si vamos a la farmacia o al almacén, al principio va bien pero, al poco tiempo, la cosa se complica. Claramente no alcanza para definirnos. Está claro que para construirnos hacen falta, al menos, dos grandes componentes: lo que traemos de fábrica, que heredamos de papá y mamá, y lo que hacemos con lo que traemos de fábrica —eso que podemos llamar ambiente, o cultura—.

Por suerte para esta construcción tenemos tiempo. Mucho tiempo. Unos miles de millones de años, por ejemplo, para entregar el informe del subsidio de investigación que nos han encargado. Y así, entre prueba y error, cambios en las condiciones de contexto, provisión de energía y temperatura adecuadas, estampitas de San Cayetano y mucha paciencia, de pronto, pasa algo. Como diría un tal doctor Víctor Frankenstein… *it's alive*. No sabemos cómo, o por qué, pero el hecho es que finalmente… sucedió .

Lo cierto es que no sabemos del todo qué es esto que somos, pero en toda definición de humano que se precie deben figurar nuestra capacidad de invención, el pensamiento abstracto, cierta fascinación

por la propiedad de las cosas, la autoconciencia, la tendencia a ser juguetones y, en algún lugar de la lista, la inteligencia.

El problema es que no entendemos muy bien qué es la inteligencia. Está bien, tenemos supuestas formas de medirla, pero no nos ponemos muy de acuerdo en qué significan estas mediciones que, por otro lado, son monstruos de laboratorio, muy distantes de lo que pueda suceder en la cancha del mundo real. Si nos atenemos a su etimología, la inteligencia tiene que ver con «saber elegir entre opciones», lo cual, si lo pensamos un poco no está nada mal: la vida es un jardín de decisiones que se bifurcan, y saber qué caminos tomar (y, sobre todo, cuáles descartar) es quizá la tarea más complicada que tenemos para enfrentar. Según esta definición, claro, los humanos no somos los únicos seres inteligentes del planeta: todos los bichos (y en alguna medida las plantas) tenemos que elegir entre distintas opciones. Quizá sí nos caracteriza el aprendizaje constante, la posibilidad de cambiar nuestras opciones a medida que transitamos más y más información, la habilidad de adaptarnos a los cambios constantes. Y en eso parece ser que sí estamos solos en el mundo. O, al menos, lo estábamos.

Como bien dice Ian McEwan en su novela *Máquinas como yo* (2019), «(los seres humanos supimos) en el otoño del siglo XX que se nos podía imitar y mejorar». Tremendo cachetazo al ego: imitar, vaya y pase pero… ¿mejorar? Sí: tanto la novela como este libro de la colección Kuaa en la editorial Vera Cartonera hablan de eso: imitar y mejorar a través de un artificio, un aparente truco de magia llamado, justamente, inteligencia artificial.[1]

Hasta hace muy poco, la posibilidad de aprender, desaprender, reaprender y cambiar sobre la marcha era algo eminentemente humano; sin embargo, como nos cuenta este libro, la IA puede entrenarse para recomendarnos opciones personalizadas, disfrazarse de médico para analizar imágenes, reconocer emociones, tomar decisiones

1 Y si vamos a la etimología, recordemos que «artificial» tiene la misma raíz que «arte», algo que requiere creatividad.

mientras nosotros estamos tirados en el sillón con el control remoto en la mano. El libro nos cuenta también algunos de los secretos para estas recetas: cómo hacer para que una computadora aprenda a aprender. Pero no se queda allí: también nos advierte sobre los aspectos éticos de la IA —detrás de un programa hay, en el fondo, un programador (o programadora), y los algoritmos son, muchas veces, nuestros pensamientos, sesgos y prejuicios hechos código.

Como con todo espectáculo, lo más interesante siempre es descubrir el truco del mago. Contrariamente a lo que podamos pensar, esto no le quita nada de maravilloso a la magia: la vuelve más humana, más comprensible, más cercana. Y ¿Aprendizaje *automágico?* hace exactamente eso: volver a la inteligencia artificial más humana, más comprensible, más cercana. No es poco.

# INTRODUCCIÓN

ENZO FERRANTE[2]

Las novelas de ciencia ficción fueron siempre una ventana al futuro. Esas historias nos han permitido imaginar cómo sería vivir en sociedades utópicas o distópicas, donde tecnologías inexistentes en la actualidad se vuelven de pronto moneda corriente. Desde Julio Verne hasta Isaac Asimov y George Orwell, la literatura nos ha invitado a imaginar futuros posibles signados por avances científico-tecnológicos que se cuelan en las sociedades hasta transformarlas en algo que nos parece ajeno, muy lejano a la vida que llevamos hoy. Sin embargo, ya no es necesario recurrir a las novelas de ciencia ficción para encontrar avances tecnológicos que han ido moldeando nuestra forma de vivir y relacionarnos. Uno de esos avances es la inteligencia artificial.

Desde hace ya más de una década, y probablemente sin que muchos de nosotros nos diéramos cuenta, la inteligencia artificial ha comenzado a permear diversas actividades de nuestra vida cotidiana.

2 Enzo Ferrante es Doctor en Informática por la Université Paris-Saclay (París, Francia) e Ingeniero de Sistemas por la UNICEN (Tandil). Realizó su postdoctorado en el Imperial College London (Londres, Reino Unido) y a fines de 2017 volvió a la Argentina como investigador repatriado al Instituto de Señales, Sistemas e Inteligencia Computacional, sinc(i) (CONICET–UNL). En el año 2020 recibió el Premio Estímulo de la Academia Nacional de Ciencias Exactas, Físicas y Naturales y el Premio Mercosur en Ciencia y Tecnología. Trabaja en el desarrollo de métodos de aprendizaje automático para el análisis de imágenes biomédicas.

Las noticias que leemos a la mañana en las redes sociales; la comida que pedimos al mediodía por el celular; el camino que tomamos con el auto cuando volvemos a casa luego del trabajo; o la serie que miramos a la noche luego de cenar: todas estas actividades están mediadas por sistemas de inteligencia artificial que nos recomiendan qué leer, comer, mirar o incluso qué camino tomar para evitar congestiones de tráfico. Pero, ¿cuándo pasó todo esto? Si las computadoras personales ya existían a finales del siglo XX, ¿por qué hemos empezado a escuchar hablar masivamente de la inteligencia artificial en nuestras vidas hace tan solo algunos años? Y más importante aún, ¿cómo funcionan estos sistemas? A lo largo de las páginas de este libro, intentaremos esbozar una respuesta para algunos de estos interrogantes.

## ALGORITMOS PARA TODXS

Para comenzar, uno de los conceptos fundamentales que necesitaremos entender es el de algoritmo. Un algoritmo es básicamente una secuencia de pasos ordenados que, al ser ejecutados, resuelven una tarea concreta. Las recetas de cocina, los instructivos para el armado de un mueble que acabamos de comprar, o los pasos que seguimos para multiplicar dos números en una hoja de papel, son ejemplos con los que interactuamos a diario, sin saber que estamos siguiendo un algoritmo. Pero en este libro, los algoritmos que nos interesan no son ni las recetas ni los instructivos de muebles, sino aquellos que pueden ser ejecutados por una computadora. Los programas de computadora, también conocidos como *software* o sistemas informáticos, son en realidad algoritmos escritos en un lenguaje particular que puede ser entendido tanto por seres humanos como por computadoras. De esta forma, las programadoras y los programadores son personas que pueden darle instrucciones a una computadora para que ejecute acciones tales como mostrar un mensaje por pantalla, sumar dos números o pedirnos que ingresemos un texto con el teclado. Combinando estas acciones es que se logran construir programas más complejos como los procesadores de texto, los

videojuegos, las planillas de cálculo o los navegadores de internet. Todos estos programas fueron escritos en un lenguaje de programación determinado que no es ni español ni inglés sino que son lenguajes con nombres como *Python, Java* o *C++*. Pero entonces ¿qué tienen que ver los algoritmos con la inteligencia artificial?

## UN VIAJE AL CORAZÓN DE LA INTELIGENCIA ARTIFICIAL CONTEMPORÁNEA

Tal como veremos en el primer capítulo de este libro de la mano de Diego Fernandez Slezak, la inteligencia artificial —entendida en un sentido amplio como la disciplina que se encarga de comprender y construir entidades artificiales inteligentes que simulan en algún sentido el comportamiento humano— comenzó a desarrollarse en los años cincuenta, mucho antes de que las computadoras personales que usamos hoy existieran como tales. En la actualidad, los algoritmos de inteligencia artificial son métodos (recetas) capaces de ser ejecutados por una computadora y que pretenden simular, en algún sentido, el comportamiento de una entidad inteligente. Esta definición de inteligencia artificial es ciertamente muy amplia, y abarca conceptos que van desde los sistemas de razonamiento deductivo basados en reglas lógicas hasta algoritmos de aprendizaje automático que buscan detectar automáticamente patrones en conjuntos de datos y luego usarlos para realizar predicciones. En este libro, nos interesaremos principalmente por este último subcampo de la inteligencia artificial: el aprendizaje automático, que ha sido el motor de una de las revoluciones más importantes de los últimos años en el campo de la computación. Revolución que se ha trasladado a todas y cada una de las disciplinas científicas y tecnológicas, impactando en nuestra vida cotidiana. Desde la biología hasta la medicina, pasando por las ciencias sociales y la física de partículas, todas estas disciplinas hacen uso hoy de métodos de aprendizaje automático.

Los algoritmos de aprendizaje automático nos permiten *entrenar* a una computadora para realizar una tarea específica (como

detectar la presencia de una persona en una imagen, o predecir si mañana lloverá) a partir del análisis de grandes bases de datos. Al utilizar estos algoritmos, la que *aprende* es la computadora. Así, los datos se transforman en la materia prima utilizada por estos algoritmos para encontrar patrones y asociaciones que nos permitan realizar las predicciones correctas. En el segundo y tercer capítulo del libro, Luciana Ferrer, Diego Milone y Georgina Stegmayer nos ayudarán a comprender cómo podemos llevar a cabo ese entrenamiento siguiendo dos paradigmas diferentes: el de aprendizaje supervisado y el de aprendizaje no supervisado. Mientras en el primero necesitamos contar con datos y su correspondiente *etiqueta* para supervisar el aprendizaje de las máquinas (etiquetas que nos indican cuál sería la respuesta correcta al problema que estamos tratando de resolver), en el segundo caso los algoritmos tratan de descubrir patrones o asociaciones sin la necesidad de contar con una referencia explícita para la supervisión.

Finalmente, el libro termina reflexionando sobre las implicancias éticas de la inteligencia artificial. ¿Son objetivos los algoritmos? ¿Podría suceder que las decisiones tomadas por un sistema automático beneficien más a un sector de la población que a otro? Y si es así, ¿cómo podemos hacer para prevenirlo? En el último capítulo del libro, Laura Alonso Alemany nos introduce al complejo mundo de la equidad algorítmica, los sesgos y la importancia de las personas en el desarrollo de estos sistemas haciéndonos comprender que aquella inocencia artificial de la que nos habla Charly García en «La máquina de ser feliz», está lejos de ser una realidad.

## ¿APRENDIZAJE AUTOMÁGICO?

Nuestro objetivo es que, al finalizar de leer este libro, escrito colectivamente por investigadoras e investigadores en ciencias de la computación de nuestro país, puedas inferir por qué su título está colocado entre signos de pregunta. Independientemente del hecho de que sea un mundo sumamente fascinante y divertido, no

hay nada de *mágico* en los algoritmos de aprendizaje automático e inteligencia artificial. La idea es que entre todas y todos podamos disipar esa aura de misterio y esoterismo que suele aparecer cada vez que alguien nombra estos conceptos. Los algoritmos poseen un trasfondo lógico matemático que los sustenta, y esperamos que con la ayuda de estos textos, puedas llegar a comprenderlo.

# UNA BREVE INTRODUCCIÓN A LA INTELIGENCIA ARTIFICIAL

DIEGO FERNANDEZ SLEZAK[3]

El término inteligencia artificial (IA) surgió en 1956 en un seminario de verano en Dartmouth College, Estados Unidos, organizado por varios investigadores que hoy son considerados como fundadores de la computación tal como la conocemos. El objetivo de ese encuentro consistió en discutir acerca del poder de cómputo de las computadoras y sobre los desafíos y problemas para utilizar esta creciente capacidad en la creación de sistemas inteligentes. Las conclusiones fueron apresuradamente optimistas, llegando a estimar que estaría resuelto el problema para la siguiente década.

Si bien las previsiones fueron erradas, de esas reuniones surgieron distintas ramas de la inteligencia artificial en el camino de la construcción de sistemas inteligentes. Esta disciplina tiene como uno de sus principales pilares la construcción de autómatas inteligentes capaces de resolver un amplio rango de problemas. Esta intención aplicada ha sido extremadamente exitosa: hoy, las torres

3 Diego Fernandez Slezak es doctor en Ciencias de la Computación, Investigador Independiente del CONICET, profesor en la Facultad de Ciencias Exactas y Naturales de la Universidad de Buenos Aires (UBA) y director del Laboratorio de Inteligencia Artificial Aplicada (LIAA, ICC, UBA). Recibió el prestigioso premio Microsoft Faculty Fellow 2014 por sus trabajos en este campo. Es, además, director de tecnología (CTO) de la empresa de base tecnológica EntelAI. Se dedica a la investigación en la frontera entre la inteligencia artificial y la neurociencia.

de control, los aviones, el tránsito y los sistemas de salud se manejan sobre premisas de la inteligencia artificial.

Sin embargo, los fundadores de la inteligencia artificial se propusieron otra gesta que resulta mucho menos próspera: generar autómatas capaces de simular la inteligencia humana, es decir que pueda camuflarse como un ser humano, con sus aciertos, con sus errores.

Entonces ¿qué es la inteligencia artificial? Para poder contestar esta pregunta, primero deberíamos preguntarnos qué es la inteligencia (¿humana?). Alan Turing, uno de los padres de la computación, hace cincuenta años se hacía la siguiente pregunta: «¿Pueden las máquinas pensar?». Una posible respuesta podría obtenerse encuestando a muchas personas acerca de esta pregunta epistemológica, pero aun así la definición resultante seguramente sería motivo de debate y discusión, sin lograr una definición de consenso sobre el pensamiento y la inteligencia. Así es que Turing propuso una solución muy astuta. No necesitamos definir inteligencia... Hablaremos de inteligencia artificial cuando una máquina pueda *replicar* el comportamiento humano a tal punto que cuando interactuemos con esa máquina no la podamos distinguir de una persona.

Muchos investigadores se han propuesto lograr desarrollar una inteligencia artificial que imite el comportamiento humano. Por ejemplo, si jugamos videojuegos por internet (como el ajedrez, juego paradigmático en el estudio de la inteligencia artificial) es probable que no nos demos cuenta si los compañeros de equipo o los oponentes son humanos jugando o es una computadora. Esta imitación incluye tanto los aciertos como los errores... podemos jugar contra máquinas que juegan muy bien o que juegan muy mal, igual que los humanos. Esto es lo que típicamente se conoce como inteligencia artificial *débil*. Débil en el sentido de que es posible replicar el comportamiento humano en dominios pequeños, acotados, pero que saliendo apenas por fuera de la tarea para la que fue concebida, queda en evidencia que lo que se encuentra del otro lado es una máquina.

Existen también muchos casos en los que la inteligencia artificial se desvía de replicar el pensamiento humano para seguir su

propio camino. Y en este camino, muchas veces ha logrado incluso superar (con frecuencia ampliamente) el rendimiento humano. El ajedrez, otra vez, podría ser un buen ejemplo. Cómo hacer que una computadora juegue bien al ajedrez ha sido motivo de estudio en el desarrollo de la inteligencia artificial desde el pionero artículo de 1949 publicado por Claude Shannon: «Programming a Computer for Playing Chess» («Programando una computadora para que juegue al ajedrez»). Desde entonces, el desarrollo de algoritmos para ganar este juego ha sido impresionante. Uno de los hitos más importante en este campo ha sido el triunfo de Deep Blue ante el campeón mundial de ajedrez Gary Kasparov entre 1996 y 1997. Deep Blue era una supercomputadora desarrollada por la empresa IBM y diseñada para jugar al ajedrez. Su principal característica era su poder de cómputo masivamente paralelo, que le daba la capacidad de evaluar una enorme cantidad de movimientos posibles en el tablero de ajedrez (lo que se conoce como *fuerza bruta*) y seleccionar la mejor jugada. Fueron varias partidas muy reñidas, donde terminó ganando Deep Blue. Algunas movidas realizadas por la computadora deslumbraron a todos. Kasparov declaró luego de haber perdido:

> *It was a wonderful and extremely human move. I had played a lot of computers but had never experienced anything like this. I could feel —I could smell— a new kind of intelligence across the table*
>
> (Fue una maravillosa y extremadamente humana movida. He jugado con muchas computadoras, pero nunca había experimentado algo como esto. Podía sentir —podía oler— que un nuevo tipo de inteligencia estaba sentada al otro lado de la mesa).

Si bien este desarrollo resultó un gran avance en el área, ha sufrido críticas respecto a su utilidad. En definitiva, Deep Blue *solo* calculaba muy rápidamente muchas posibles movidas. En este sentido, nuevamente IBM decidió plantear otro desafío: una computadora que fuera capaz de ganarle a un humano en el juego Jeopardy. Este juego

consiste en responder preguntas de conocimiento general, como por ejemplo: ¿Cuál es la capital de Sri Lanka? Para ello, la computadora contaría con conexión a Internet en donde podría consultar la información necesaria para contestar la pregunta. En este nuevo escenario, el desafío ya no era (solamente) calcular muy rápido, sino también decodificar la pregunta, buscar los datos en la web, y luego elaborar una respuesta.

En este caso, decodificar la pregunta propone muchos problemas a resolver. Por un lado, debe interpretarse el audio con la pregunta para su transcripción. Luego, la computadora debe poder *leer* esta pregunta para elaborar una consulta en su virtualmente infinita base de datos: internet. Una vez decodificada la pregunta, se deben interpretar los resultados. Es conocido que prácticamente cualquier búsqueda en la web genera millones de resultados en los buscadores. Elegir cuál de ellos es el mejor para la elaboración de la respuesta no resulta un desafío menor. En 2011, el sistema Watson creado por IBM le ganó al campeón mundial de Jeopardy por una diferencia abrumadora.

Esta evolución tecnológica se basa tanto en el avance del poder de cómputo como en los algoritmos utilizados. El sitio web TOP500 reúne información sobre las supercomputadoras instaladas alrededor del mundo y realiza una competencia para definir cuál es la supercomputadora más veloz del planeta. En el año 1996, cuando Deep Blue ganó la primera partida a Kasparov, la computadora más veloz tenía una potencia de 368,2 GFLOPS, es decir más de 300 mil millones de operaciones de números con coma (sumas, restas, multiplicaciones, divisiones) por segundo. En el *ranking* de 2016, la computadora más potente proveía más de 93 PFLOPS, es decir un millón de veces más que en 1996.

Pero entonces, ¿qué es lo que hace tan poderosas (acaso, inteligentes) a las máquinas? En primer lugar, como mencionamos, la evolución en la velocidad de cómputo matemático. Y, a su vez, esto es acompañado de su memoria prácticamente infinita. Pero la verdadera revolución ha tenido lugar en la última década. El pilar sobre el que se sostiene el éxito de la inteligencia artificial contemporánea

son los programas desarrollados en los últimos tiempos que permiten identificar patrones, repeticiones, parecidos y diferencias en grandes volúmenes de datos. Esto es lo que algunos tal vez hayan escuchado nombrar como aprendizaje automático, o *machine learning*: métodos computacionales capaces de aprender a sacar conclusiones a partir de una gran cantidad de datos. En particular, la gran revolución de la última década ha sido protagonizada por una técnica específica de aprendizaje automático conocida como aprendizaje profundo, o *deep learning*. Contrariamente a lo que se hacía previamente, cuando expertos indicaban las reglas que debían seguir las computadoras para extraer los resultados, estos nuevos algoritmos infieren sin ayuda de expertos (más que los datos de entrada a partir de los cuales se hace el análisis) las reglas y patrones asociados.

Hasta hace poco decíamos que la computadora se limitaba estrictamente a hacer aquello para lo que la programamos, como si estuviera siguiendo una receta de cocina. Pero en el último tiempo, esta idea cambió. La capacidad de encontrar patrones o repeticiones y de acumular información permite que programas aprendan a partir de los datos y lleguen a conclusiones, a resultados, sin tener que programarlos explícitamente. Es decir que fueron inferidos, aprendidos, justamente a partir de la repetición en la enorme disponibilidad de datos.

Pero este es solo uno de los subcampos de la inteligencia artificial, conocido como aprendizaje automático. Volviendo a la definición original surgida del seminario en Dartmouth College, esta sería una de las tantas ramas de la inteligencia artificial a resolver: la del aprendizaje. Y será la que discutiremos en los siguientes capítulos de este libro. Existen, sin embargo, otras áreas de mucho desarrollo como los sistemas expertos y los de razonamiento lógico que no serán abordadas. Hoy, prácticamente todos los sistemas que dicen utilizar inteligencia artificial, en realidad, incorporan solo alguna de estas ramas, típicamente la de aprendizaje automático.

El camino hacia una inteligencia artificial general —una máquina inteligente capaz de replicar el comportamiento a tal

punto de que no seamos capaces de darnos cuenta si estamos interactuando con una computadora o un ser humano— aún es largo y queda mucho por recorrer. Hoy los desarrollos en el área permiten explorar nuevas fronteras del conocimiento en una sinergia entre inteligencias humanas combinadas con inteligencias artificiales, lo que muchos gustan llamar una inteligencia aumentada.

# APRENDIZAJE SUPERVISADO, O CÓMO ENTRENAR A TU COMPUTADORA

LUCIANA FERRER[4]

Desde que nacemos, gran parte de nuestro aprendizaje sucede sin supervisión. Nadie nos enseña que las cosas, si las soltás en el aire, caen al piso. Lo aprendemos de bebés observando una y otra vez qué pasa cuando tiramos algo al aire (un ejercicio que los bebés disfrutan muchísimo). Tampoco nos enseñan, en general, que para gatear necesitamos mover las extremidades de una cierta manera. Aprendemos a desplazarnos por prueba y error.

Sin embargo, también hay muchas cosas que aprendemos con supervisión de la gente que nos rodea. Por ejemplo, muchas palabras las aprendemos porque alguien de vez en cuando señala u ofrece algo y dice su nombre: «¿Querés agua?» o «mirá qué lindo perrito», o «eso es rojo». Más adelante, a medida que crecemos, aprendemos muchas más cosas de manera supervisada: a leer y escribir, sumar, restar, multiplicar, geometría, qué es un verbo, cómo agarrar

4 Luciana Ferrer es Investigadora Adjunta en el Instituto de Ciencias de la Computación, UBA–CONICET. Completó su grado de Ingeniería Electrónica en la Universidad de Buenos Aires en 2001 y su doctorado en Ingeniería Electrónica en Stanford University (California, Estados Unidos) en 2009. Su tema principal de investigación es el aprendizaje automático aplicado al procesamiento del habla. Sus investigaciones abordan problemáticas como el reconocimiento de la identidad o el estado mental de un hablante a partir de su voz, la identificación de idiomas en grabaciones de habla, o la calificación de la calidad de la pronunciación de un estudiante de inglés.

*bien* el tenedor, cómo atarnos los cordones. Todo eso lo aprendemos con ayuda, con el ejemplo o con instrucciones, paso a paso.

Así como los humanos, los sistemas de aprendizaje automático también pueden aprender muchas cosas de manera supervisada. En la actualidad estamos rodeados de sistemas de inteligencia artificial que aprendieron a resolver problemas de esta manera. Por ejemplo, los sistemas que reconocen lo que estás diciendo y lo transcriben a palabras, fueron generados usando miles de horas de grabaciones de personas hablando, acompañadas por sus correspondientes transcripciones (hechas a mano por humanos). Otro ejemplo es el de detección de *email* basura (*spam*). Este algoritmo se entrena a medida que los usuarios del servicio de *email* van anotando sus correos como basura. Usando esos datos, anotados por humanos, se puede desarrollar un sistema que hace la predicción de manera automática y le evita al usuario ver decenas o cientos de *emails* por día intentando venderle diversos productos que la persona no necesita.

Los sistemas de auto-corrección, o de predicción de la próxima palabra que vas a tipear, también se aprenden de manera supervisada. En ese caso, a veces se los llama auto-supervisados, porque no hace falta que nadie anote los datos, simplemente se usa texto generado por humanos, pero no anotado especialmente para resolver este problema. Dado mucho texto, millones y millones de palabras escritas, los algoritmos aprenden cuáles son las palabras o secuencias de palabras más frecuentes en un idioma, y usan esa información para corregirte cuando metiste mal el dedo en el teclado, o para predecir la próxima palabra y ahorrarte el trabajo de tipearla.

Finalmente, otro ejemplo que vamos a seguir usando el resto de este capítulo es el de detección de emociones. Si queremos saber cuándo una persona que está hablando se encuentra enojada, triste o alegre, necesitamos que el sistema aprenda cómo suena el enojo, la tristeza y la alegría. Para eso le damos un montón de grabaciones (a las que llamaremos *muestras*) donde una o varias personas anotaron qué emoción escuchan en cada caso. Esas anotaciones realizadas por personas las llamaremos *etiquetas*. Con esas grabaciones

y sus correspondientes etiquetas, podemos desarrollar un sistema que, luego, pueda predecir de manera automática la emoción de una grabación con habla.

La pregunta que resta responder es, entonces, ¿cómo se construyen estos sistemas? Los sistemas de clasificación supervisada, en general, están compuestos por una etapa de extracción de características para cada muestra, seguida de una etapa de modelado de estas características que es la que genera las predicciones. El modelo se aprende usando datos anotados con la etiqueta que queremos predecir. Una vez aprendido el modelo, generalmente se evalúa su rendimiento, es decir, cuán bien o mal funciona en otros datos. A continuación, describimos con más detalle estos conceptos.

## LAS CARACTERÍSTICAS

Para poder generar un programa de computadora que clasifique datos de manera automática, necesitamos representar estos datos, cada una de las muestras, como un conjunto de números. La representación que elijamos va a depender del tipo de datos (texto, audio, imágenes, etc.), de la tarea a resolver y del modelo que pensamos usar. Por ejemplo, para hacer reconocimiento de emociones, hay ciertos aspectos de la señal de habla que son importantes: la intensidad (el volumen), la velocidad (medida, por ejemplo, como la cantidad de sílabas pronunciadas por minuto) y el tono del habla (agudo o grave) y sus variaciones en el tiempo. Alguien que habla muy monótonamente, sin variar el tono o la intensidad, es poco probable que esté enojado o contento (aunque depende un montón de la persona, lo cual complica mucho las cosas). Por lo tanto, de acuerdo con esta intuición, muchos de los sistemas de reconocimiento de emociones usan como características un conjunto de medidas basadas en la intensidad, el tono y la velocidad. Por ejemplo, un conjunto mínimo de características podría incluir la intensidad, tono y velocidad promedio a lo largo del tiempo, junto con alguna medida de su variación (como el rango, calculado como la diferencia entre

el máximo y el mínimo). Estos valores, concatenados uno detrás del otro en lo que se conoce como un vector (algo así como una fila de una tabla en Excel, donde cada columna representa una característica), serán las características de las muestras, tal como lo muestra el siguiente esquema:

| | RANGO DE VARIACIÓN DE LA INTENSIDAD | RANGO DE VARIACIÓN DEL TONO | VELOCIDAD PROMEDIO | ETC... |
|---|---|---|---|---|
| MUESTRA 1 | 0,32 | 300,2 | 4,3 | ... |
| MUESTRA 2 | 0,43 | 430,1 | 2,5 | ... |
| ... | ... | ... | ... | ... |

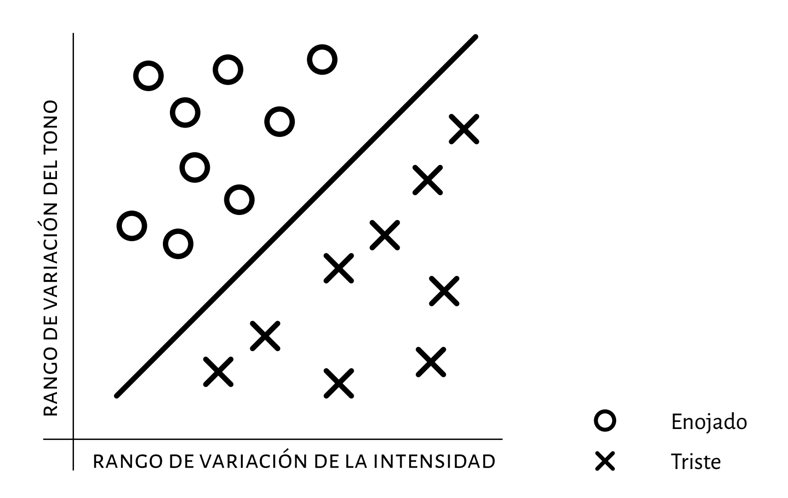
En tareas de procesamiento de imágenes (por ejemplo, para encontrar qué objetos aparecen en una foto), las características muchas veces son directamente las intensidades de cada color (rojo, verde y azul) asociadas a cada pixel. En tareas de procesamiento del lenguaje, las palabras pueden ser representadas como vectores enormes del tamaño del vocabulario completo que se está considerando (en castellano, por ejemplo, la Real Academia Española lista casi 100 mil palabras). El vector correspondiente a cada palabra contiene ceros en todos lados, salvo en el lugar correspondiente a esa palabra, donde se coloca un uno.

Sea cual sea la tarea o el tipo de datos a procesar, siempre necesitamos representar esos datos de manera numérica en vectores de características que luego usamos como entrada al modelo.

## EL MODELO

Imaginemos un caso muy sencillo en el que tenemos solo dos características por cada muestra y solo dos posibles etiquetas que

podremos asignarles. Por ejemplo, en el caso de la clasificación de emociones, las posibles etiquetas podrían ser «enojado» o «triste», y las dos características que describen a cada muestra podrían ser: (1) rango de variación de la intensidad y (2) rango de variación del tono. Claramente, esos dos datos no alcanzan para clasificar emociones, pero asumamos que sí. En ese caso, podríamos dibujar nuestras muestras como círculos y cruces en un plano, donde los ejes X e Y son las dos características, y la forma utilizada (círculo o cruz) indica la clase a la que pertenecen, es decir, la etiqueta asignada por los anotadores, tal como se ve en la figura.




En este caso súper simplificado, las muestras etiquetadas como «enojado» se pueden separar de las muestras etiquetadas como «triste» con una recta. Esa recta está definida por la siguiente forma matemática $y = a . x + b$, donde $a$ y $b$ son los parámetros de la recta y representan a nuestro modelo. Es decir, una vez obtenidos los parámetros $a$ y $b$, vamos a poder asignarle una etiqueta a cualquier nueva muestra (representada por su valores $x$, el rango de variación de la intensidad, e $y$, el rango de variación del tono); conceptualmente

solo tenemos que dibujar la recta y ver de qué lado cae la muestra, del lado de los círculos o del lado de los triángulos.

Más generalmente, el modelo es una expresión matemática que describe el comportamiento de nuestros datos. Para obtener el modelo necesitamos empezar planteando ciertos supuestos y combinarlos con la información presente en nuestros datos. Por ejemplo, en el caso de la figura anterior, nuestro supuesto podría haber sido: «una recta debería poder separar bien las dos clases, dadas estas dos características». Una vez planteado este supuesto, nos falta encontrar los parámetros de esa recta. Para eso, usamos los datos anotados disponibles. A ojo, en este caso simple podríamos dibujar la recta que los separa mejor y calcular sus parámetros. En la práctica, sin embargo, rara vez tenemos solo dos o tres características y, lamentablemente, no podemos visualizar datos en más de 3 dimensiones. Por lo tanto, necesitamos un procedimiento para que, de manera automática, podamos encontrar los mejores parámetros posibles para esa recta que suponemos puede funcionar bien para resolver nuestro problema.

La manera de encontrar estos valores consiste en plantear una función matemática que mida cuán bueno o malo es un cierto conjunto de valores para los parámetros, dados nuestros datos de entrenamiento. La llamamos *función de costo*: cuanto más chico sea el valor de esta función, mejor. Como estamos hablando de aprendizaje supervisado, estos datos de entrenamiento incluyen no solo las características para cada muestra, sino también su etiqueta (en nuestro ejemplo, la emoción). Dada la función y los datos, lo que hacemos a continuación es buscar el conjunto de parámetros que resulta en el valor más chico (llamado *mínimo*) para esa función. El conjunto de parámetros obtenido será el que representa nuestro modelo. Hay muchas opciones para definir la función de costo, dependiendo del problema, de la cantidad de datos disponibles y de la preferencia de la persona a cargo de desarrollar el sistema.

Por ejemplo, un método tradicional de aprendizaje supervisado es la máquina de soporte vectorial (sí, un nombre un poco grandilocuente). En su versión más simple, los SVMs (por sus siglas

en inglés, *Support Vector Machines*) asumen una recta (o su versión para más de dos dimensiones que se llama *hiperplano*) para separar las dos clases, como en nuestro ejemplo anterior. Pero, a diferencia de nuestro caso donde encontramos la recta a ojo, los SVMs usan una función de costo para encontrar los parámetros de la recta. La función de costo está dada por una expresión matemática que pide que todos los puntos de cada clase estén lo más lejos posible de la recta, del lado correcto, claro. O sea, que el margen entre la recta y el punto más cercano de cada lado sea lo más grande posible. La recta que dibujamos a ojo en la figura anterior, más o menos, parece cumplir este requisito.

Ahora, ¿qué pasa si no es posible dibujar una recta que separe las dos clases completamente porque las muestras de ambas clases se mezclan en el borde? ¡En la práctica este es el caso más común! Rara vez las clases son perfectamente separables por una recta. Para considerar ese caso, los SVMs relajan un poco la restricción y permiten que algunos puntos estén más cerca de la recta o incluso del lado incorrecto, pero para cada punto en el que pasa esto, se agrega una *penalización* a la función de costo. La meta es, entonces, encontrar la recta, con el menor costo total.

Los modelos descriptos arriba son las versiones más sencillas de clasificadores binarios (porque hay dos clases a distinguir), pero hay muchas otras opciones que asumen modelos más complejos o que sirven para clasificar en más de dos clases o incluso para aprender a predecir un valor continuo (como podría ser el precio que va a tener un cierto producto en el futuro).

## EVALUACIÓN DEL MODELO

Una vez que contamos con un modelo entrenado, es decir, cuando tenemos su expresión matemática completa, debemos evaluar cuán bien funciona para el objetivo planteado. Esto es importante, por ejemplo, para saber si el sistema es suficientemente bueno para nuestros propósitos o para elegir el mejor entre dos modelos posibles. Para esta evaluación consideramos muestras que no hayan

sido usadas para entrenar el sistema ni para tomar decisiones durante el desarrollo porque esto llevaría a resultados demasiado optimistas: el sistema seguramente se va a comportar mejor en los datos que usamos para entrenar y tomar decisiones de desarrollo que en datos nuevos. Además, idealmente, esas muestras tienen que ser representativas del uso final que le vamos a dar al sistema. Si queremos usar el sistema para detectar la emoción de un cliente cuando llama al negocio para quejarse porque algo que le vendieron no le funciona bien, necesitamos ese tipo de datos. No podemos usar muestras de gente conversando con su familia porque muy probablemente el sistema funcione bastante distinto en ambas situaciones.

Una vez que tenemos los datos, lo que debemos hacer es aplicarles el sistema. Es decir, generar las respuestas para cada una de esas muestras. Para evaluar si el sistema funcionó bien o no, tenemos que comparar esas respuestas con la etiqueta real de cada muestra. La manera más sencilla de medir cuán bien le fue al sistema en estos casos es calcular la proporción de muestras en las que el sistema detectó la clase correcta. A esto le llamamos la *exactitud* del sistema. Idealmente querríamos un sistema con una exactitud del 100%, pero eso rara vez sucede, y generalmente nos tenemos que conformar con exactitudes menores (a veces mucho menores) al 100%.

En resumen, los algoritmos de aprendizaje supervisado buscan aprender patrones a partir de datos de entrenamiento donde para cada muestra tenemos anotada la etiqueta o clase a predecir. El proceso de aprender estos patrones, en general, consiste en dos pasos principales: extracción de características relevantes para el problema y modelado usando una serie de supuestos sobre la relación entre esas características y la clase a predecir. El modelo está compuesto por una expresión matemática que incluye parámetros. Estos parámetros los encontramos usando datos etiquetados con la clase de interés. También usamos datos anotados para ver si el modelo que aprendimos efectivamente es capaz de predecir correctamente lo que esperamos, es decir, lo evaluamos para conocer la *exactitud* del sistema.

# CUANDO LAS COMPUTADORAS APRENDEN SOLAS

DIEGO MILONE[5]
GEORGINA STEGMAYER[6]

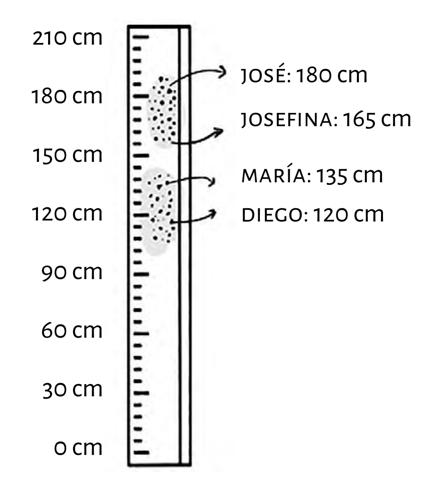
En el capítulo anterior, vimos ejemplos de algoritmos de aprendizaje supervisado, que son capaces de aprender a realizar tareas específicas a partir de datos de entrenamiento con etiquetas que fueron provistas por humanos. Ahora bien, ¿es posible que las computadoras aprendan por sí mismas, sin ningún tipo de supervisión humana?

Imaginemos que un día, en una escuela, alguien se puso a medir estaturas. Al final del día había una larga lista que tenía la altura en centímetros y los nombres de cada persona medida. Había que analizar toda esa información. Como eran muchas mediciones, se fueron dibujando puntitos sobre un centímetro: cada altura con

5 Diego Milone es Investigador Principal del CONICET y Profesor Titular de la Universidad Nacional del Litoral. Fue fundador y director (hasta el año 2020) del Instituto de Investigación en Señales, Sistemas e Inteligencia Computacional, sinc(i), de la Universidad Nacional del Litoral y el CONICET. Recibió en 2009 el Premio Dan J. Beninson de la Academia Nacional de Ciencias Exactas, Físicas y Naturales, el Premio Sadosky de Oro de la Fundación Sadosky en 2012 y el Premio Houssay en la edición 2013 de la Distinción Investigador/a de la Nación del MinCyT. Es especialista en inteligencia artificial y sus aplicaciones en análisis de señales biomédicas y bioinformática.

6 Georgina Stegmayer es Investigadora Independiente del CONICET en el Instituto de Investigación en Señales, Sistemas e Inteligencia Computacional, sinc(i). Completó su grado en Ingeniería en Sistemas de Información en la Universidad Tecnológica Nacional – Facultad Regional Santa Fe y su doctorado en el Politécnico di Torino (Italia). Es especialista en inteligencia artificial y aprendizaje automático aplicados a la bioinformática, es decir al análisis e interpretación de datos biológicos.

su nombre. ¡Vean cómo quedó el centímetro!



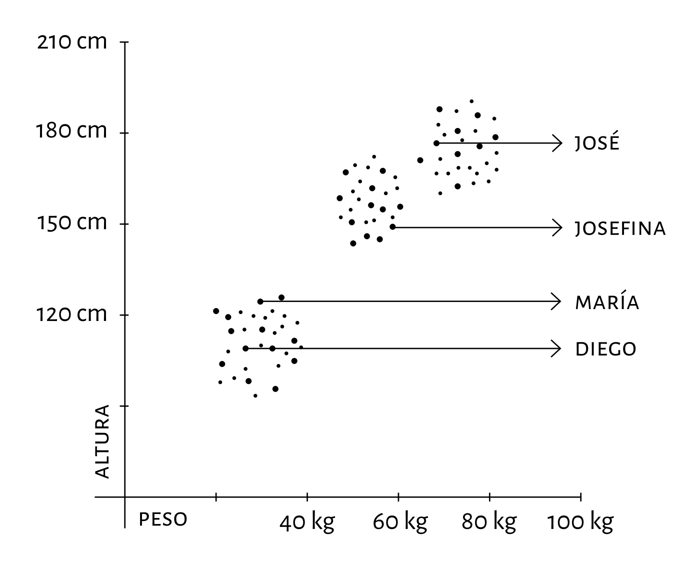

Si miramos con detalle la figura, podemos empezar a descubrir algunas cosas interesantes. Enseguida detectamos 2 *grupos* y ahí nos damos cuenta de que en la escuela habían medido a chicos y a grandes. Pero hay todavía más cosas interesantes en esos grupos de puntitos. Podemos ver también que faltaban puntos entre 140 y 160 cm. ¿Qué nos dice eso? Probablemente que es una escuela primaria, porque si bien no sabemos las estaturas para cada edad, nos podemos imaginar que si faltan alturas entre chicos y grandes, es porque no hay nadie entre 13 y 18 años.

Tal como ilustra el ejemplo, son muchas las cosas que se pueden descubrir a partir de unos grupos de puntitos. Pero, ¿qué más podemos ver? En cada grupo vemos que los puntos se acumulan y quizás podríamos pensar en que hay una estatura típica o propia de cada grupo, como si fuera el *representante* o el *centro* de cada grupo. ¿Cómo podríamos hacer para encontrar ese centro que lo caracteriza? Si hubiera solo 2 mediciones de alturas en el grupo, por ejemplo 160 cm y 180 cm, podríamos hacer 160+180 dividido por 2 y así obtendríamos la altura media (o promedio) que en este caso sería 170 cm. Si tuviéramos 3 alturas haríamos 160+180+173 dividido 3 y nos daría 171 cm. Siguiendo de esta forma, si tenemos muchos puntos en un grupo, los sumamos y dividimos por su cantidad. Ese número es el *centro* que podríamos decir que resume o que caracteriza lo que tiene el grupo.

Lo mismo podríamos hacer para el otro grupo, ¿no? Vamos a suponer que hacemos las cuentas y nos da 131 cm. Así ahora cada grupo tendría su centro que lo represente. ¿Y para qué puede servir eso? Supongamos que se miden nuevas alturas y queremos sa-

ber si son del grupo de grandes o del grupo de chicos. Imaginemos que alguien mide 178 cm. En ese caso es sencillo inferir que con esa altura pertenece al grupo de grandes. Pero supongamos que llega alguien con 158 cm, ¿en qué grupo estaría? Podemos poner el punto en el centímetro y ver si queda más cerca de un grupo que del otro. Eso que hacemos *a ojo* también lo podríamos hacer calculando la *distancia* que hay entre la nueva altura y algún punto de cada grupo, y así el que esté más cerca, gana. ¿Pero qué punto elegimos en cada grupo? Una idea podría ser elegir el más cercano, pero habría que medir la distancia a todos para saber cuál es el más cercano. Otra idea sería usar el *centro* de cada grupo. Anteriormente dijimos que el centro era como un resumen del grupo, así que nos puede venir bien. Lo que nos estaría faltando ahora es poder medir la distancia. Tenemos la nueva altura de 158 cm, y por ejemplo tenemos el centro de grandes en 169 cm y el de chicos en 125 cm. Entonces, ¿cuál está más cerca? ¿Cómo podemos hacer para descubrirlo? Una idea simple: restando. Hacemos 171–158 = 13 cm y 158–125 = 33 cm. No hay dudas. ¡Esta persona de 158 cm es del grupo de grandes!

¿Qué pasaría ahora si además de las alturas, se preguntan los pesos de cada persona? La lista en papel se ve casi igual pero con dos medidas por persona. Podríamos dibujar también los pesos en otra línea como la del centímetro y encontrar grupos. Pero hay situaciones que no son tan fáciles de ver cuando hay muchos puntos. Por ejemplo, ¿podría haber una persona muy delgada en el grupo de grandes? Entonces para poder ver todas las medidas a la vez ponemos dos líneas, una vertical con la altura y otra horizontal con el peso. Luego, si Marita mide 170 cm y pesa 65 kg, ponemos un punto en esas coordenadas como muestra el dibujo. De esta forma incluimos todos los puntitos y nuevamente nos disponemos a observar. Para nuestra sorpresa, parece que no había solo dos grupos. ¡Ahora vemos que hay 3 grupos! Esto lo encontramos porque pudimos ver los datos en 2 *dimensiones* a la vez. Si viéramos los datos en 1 dimensión (como vimos las alturas) y luego la otra dimensión




por separado (con los pesos), nunca hubiéramos encontrado este nuevo grupo. ¡Y vaya que hicimos un gran descubrimiento! Puede ser clave para un plan de salud o si estuviéramos pensando en un comedor para la escuela, que hiciera dietas diferenciadas según las necesidades. Esto de haber descubierto que hay 3 grupos sería de gran utilidad.

Como antes, quisiéramos tener un *centro* que resuma a cada grupo, pero ahora para ubicar estos centros necesitamos 2 valores, porque estamos en 2 dimensiones: altura y peso. Así que cada centro tendría una altura y un peso. ¿Cómo los calculamos? Bueno, fácil, sumando cada cosa por separado y dividiendo por la cantidad de personas en el grupo. Sumamos todas las alturas de un grupo (igual que en 1 dimensión) y la dividimos por la cantidad de personas en ese grupo. Ya tenemos la primera coordenada del centro. Luego sumamos los pesos del grupo y dividimos también por la cantidad de personas en ese grupo. ¡Y listo! Tenemos las dos coordenadas para el centro de ese grupo. Lo mismo podemos hacer para los otros 2 grupos. Supongamos que las cuentas nos dan lo siguiente:

— Centro del grupo 1: 171 cm y 68 kg
— Centro del grupo 2: 125 cm y 35 kg
— Centro del grupo 3: 162 cm y 50 kg

Y ahora, ¿cómo hacemos con las nuevas personas que se van midiendo? ¿A qué grupo las asignamos? Como hicimos antes, podríamos medir la *distancia* al centro de cada grupo y la asignamos al grupo más cercano. Pero ahora tenemos 2 números, altura y peso, y los centros también tienen 2 números. Entonces la distancia debe tener en cuenta ambas dimensiones. Una forma simple de hacerlo es sumando las distancias por separado. Por ejemplo, si la nueva medida es de 158 cm y 55 kg:

— Al centro del grupo 1: |158–171| = 13 cm y |55–68| = 13 kg, distancia total 13+13 = 26
— Al centro del grupo 2: |158–125| = 33 cm y |55–35| = 20 kg, distancia total 33+20 = 53
— Al centro del grupo 3: |158–162| = 4 cm y |55–50| = 5 kg, distancia total 4+5 = 9

Así, la nueva persona estaría en el 3er grupo.

Como vimos, tener más información de cada persona nos sirvió para descubrir nuevos grupos, y eso resultó muy interesante y útil. Entonces, si agregamos más medidas podríamos realizar más descubrimientos impensados. Incluyendo las edades, ya estamos en 3 dimensiones. Dibujarlo se puede complicar un poco más, pero las cuentas para encontrar los centros y medir distancias siguen siendo las mismas, siempre con cada medida por separado, sumando, dividiendo y restando. Y si agregamos otra medida más, por ejemplo la distancia desde la escuela hasta su casa, ¡ahora tenemos 4 dimensiones! En ese caso, ya no podemos hacer ningún dibujo, ni siquiera imaginarnos eso. Pero podemos calcular los centros y medir distancias para saber a qué grupo pertenece cada persona. Al fin y al cabo, son sumas, restas y divisiones, pero en muchas dimensiones. Pensemos, ¿cómo hacemos con 100 dimensiones? En

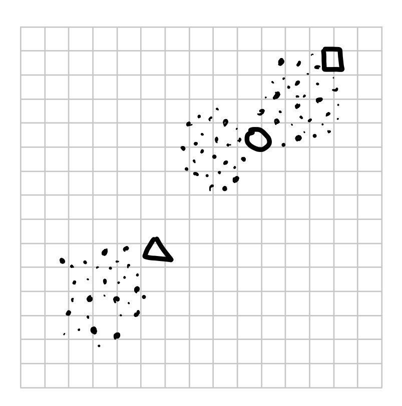
ese caso, ya se empieza a complicar hacerlo a mano: necesitamos una computadora.

Con la computadora vamos a poder agregar cada vez más información de las personas y descubrir más grupos interesantes, según la música que les gusta, sus perfiles en las redes sociales, los deportes que practican, y muchas dimensiones más. Pero hay algo que se torna complicado cuando tenemos muchas dimensiones: ahora no podemos *ver* los grupos de puntos. Tenemos que encontrar la forma de descubrir los grupos cuando hay muchos datos y cada uno con tantas dimensiones que no podemos verlas simplemente en un dibujo.

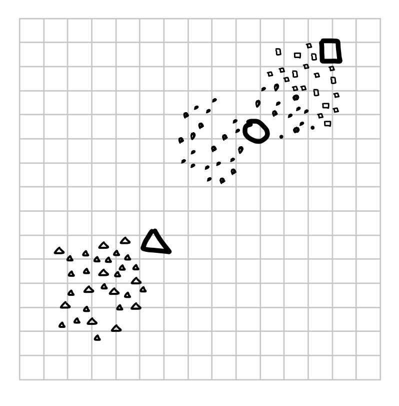
¿Es posible entonces que una computadora aprenda a descubrir esos grupos ocultos en los datos sin nuestra ayuda? La respuesta es sí: la computadora puede aprender a descubrir grupos o relaciones *ocultas* siguiendo los pasos de un algoritmo de aprendizaje no supervisado. El algoritmo más famoso para descubrir relaciones ocultas o grupos ocultos entre datos, sin ningún tipo de supervisión por parte del ser humano, se llama *k–medias*. A diferencia de los algoritmos que estudiamos en el capítulo anterior, se dice que este algoritmo es *no supervisado* porque no sabe de antemano lo que va a encontrar: lo descubre solo, sin supervisión. Es decir, no hay un ser humano que le diga qué es lo que tiene que encontrar en los datos. En este caso los datos no tienen etiquetas que indiquen de qué grupo son y el algoritmo tiene que aprender a identificarlos. Uno le dice al algoritmo «tenés que encontrar k grupos entre los datos» (podría ser k=2 o k=3 grupos, por ejemplo) y la computadora los descubre, sola. Lo que hace este algoritmo es mirar los valores en las dimensiones de cada dato, e ir encontrando los que son muy parecidos entre sí, y a la vez muy diferentes de los otros. Como vimos en el ejemplo de la escuela, las dimensiones de cada dato son números que describen o caracterizan a cada persona, como su altura, peso, edad, distancia de su casa a la escuela, etc.

El algoritmo *k–medias* recibe los datos de entrada y la cantidad *k* de grupos que tiene que encontrar. Luego sigue unos pasos muy simples:

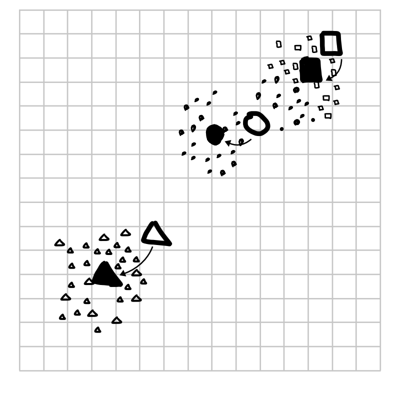
**1) Inicio:** se eligen *k* datos cualquiera como centros de los grupos (marcados con símbolos más grandes en el diagrama). Supongamos que estos centros son centro1 (cuadrado), centro2 (círculo) y centro3 (triángulo). Para las personas medidas en la escuela (puntos negros), decidimos encontrar 3 grupos, así que k=3. En esta instancia, no nos preocupa dónde están estos centros iniciales, luego el algoritmo los va a acomodar.



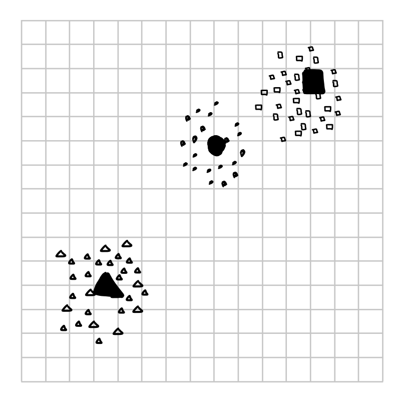
**2) Asignación de datos a grupos:** se va mirando cada dato y midiendo la distancia a cada centro como: *distancia = |dato - centro|* Cada dato se marca como perteneciente al grupo de su centro más cercano. En la figura los puntos negros que se convirtieron en cuadrados se asignaron al centro1, los pequeños círculos al centro2, y los triángulos al centro3.



**3) Actualización de los centros:** cada vez que se asignan datos a los centros, los grupos cambian y los centros ya no representan bien el punto medio de cada nube de puntos. Así que hay que volver a definir los centros, como hicimos antes con las alturas y pesos. Entonces ahora cada centro se calcula como el promedio de todos los datos que están en su grupo. En la figura, los nuevos centros se representan como símbolos grandes rellenos de negro.



**4) Se vuelve al paso 2:** se repiten los pasos 2 y 3, muchas veces, hasta que los centros ya no cambian de lugar. Cuando los centros quedan fijos, el algoritmo ha descubierto los *k* grupos que le pedimos que encuentre en los datos.



Al finalizar el algoritmo de *k-medias*, podemos concluir que la computadora *aprendió* de estos datos, y lo hizo sola, por lo que decimos que aprendió la estructura oculta, la fue descubriendo. Como resultado de ese aprendizaje *no supervisado* nos quedaron los datos agrupados según sus características y obtuvimos también *k* representantes de cada grupo, los centros, que tienen las mismas dimensiones que los datos (altura, peso, etc) y para cada una nos dice cuál es el valor que caracteriza a ese grupo de personas.

Los grupos que se encuentren con este algoritmo pueden confirmar alguna teoría o idea inicial que teníamos sobre los datos. Y también puede ayudarnos a descubrir relaciones asombrosas entre los datos, que si son muchos o tienen muchas dimensiones, de manera manual o mirando *a ojo* no hubiéramos reconocido jamás. Imaginemos millones de datos de los *tweets* de millones de usuarios en la web. Cada usuario puede ser descripto por sus datos personales (edad, profesión, ubicación geográfica, cantidad de seguidores) y por los tipos de *tweets* que hace (cantidad de palabras promedio, tipo de palabras, cantidad de tweets al día, *likes*). Todas estas son las dimensiones características de cada usuario. Ahora supongamos que queremos encontrar grupos de usuarios con un perfil en común. Esto es imposible para una persona e incluso para miles de personas mirando los datos. Pero una computadora podría analizar toda esta información en segundos con k–medias y descubrir, de forma no supervisada, los principales perfiles de usuarios en la red. Así funcionan los algoritmos de aprendizaje no supervisado.

# INTELIGENCIA ARTIFICIAL Y VALORES

LAURA ALONSO ALEMANY[7]

## LOS SISTEMAS AUTOMÁTICOS ¿SON OBJETIVOS?

En los capítulos anteriores hemos visto la inteligencia artificial como una tecnología que automatiza sistematicidades, ofrece predicciones a partir de ejemplos o descubre patrones en los datos. Los resultados pueden no ser perfectos, pero podemos medir el error de estos programas con métricas estándares, bien establecidas. Todo esto nos lleva a pensar que se trata de una tecnología más objetiva que los seres humanos, seres subjetivos y prejuiciosos. Nuestra creencia es que los sistemas basados en datos son especialmente objetivos porque ni siquiera incorporan la subjetividad del programador que escribe unas reglas según sus intuiciones, sino que se construyen enteramente a partir de datos *objetivos*.

Sin embargo, los sistemas de inteligencia artificial, incluso los basados en datos, incorporan las subjetividades de los equipos que los crean y de los grupos sociales que los financian. Esas

7 Laura Alonso Alemany es doctora en Lingüística Computacional por la Universitat de Barcelona, profesora e investigadora en la Universidad Nacional de Córdoba y encabeza el grupo de investigación en Procesamiento de Lenguaje Natural de la Facultad de Matemática, Astronomía, Física y Computación en dicha universidad. Es especialista en inteligencia artificial aplicada al procesamiento del lenguaje.

subjetividades pueden llegar a resultar perjudiciales para parte de la población, incluso de formas muy sutiles, como veremos a continuación.

## LOS EFECTOS VAN MÁS ALLÁ DE LOS RESULTADOS

Hemos visto que el error forma parte de los sistemas de inteligencia artificial, y lo hemos aceptado e integrado en nuestra convivencia con estas tecnologías. Cuando pensamos en estos sistemas, entendemos que pueden tener algunas limitaciones. Muchas veces tratamos de adaptar nuestro comportamiento para obtener buenos resultados, como por ejemplo cuando pronunciamos *bien* para que un reconocedor de voz identifique correctamente las palabras que queremos comunicar.

Detengámonos un momento en esta escena. ¿Qué implica que *pronunciemos bien*? En muchos casos, no solamente implica que vamos a tratar de ser claros, sino que también vamos a adaptar nuestra forma de hablar a lo que sabemos que la máquina reconoce. Y ¿qué reconoce la máquina? Como vimos en los capítulos anteriores, el sistema reconoce lo que aprendió a reconocer a partir de ejemplos. Pero esos ejemplos, ¿de dónde salen?

El castellano tiene muchas variantes, algunas tan distintas que la comprensión entre hablantes de diferentes variantes resulta prácticamente imposible. Los hablantes del castellano, como también los de otras lenguas con variantes muy distintas, como el alemán, italiano o inglés, muchas veces consiguen entenderse entre sí porque aprendieron, además de la variante que es su lengua materna, una variante llamada *estándar* que facilita la intercomprensión entre hablantes. ¿Cómo se establece cuál es la variante estándar? Por lo general, se trata de la variante de un grupo dominante, como por ejemplo el castellano de Castilla (la lengua de los conquistadores) o el latinoamericano neutro (la variante del castellano que eligen los grandes medios de comunicación internacionales). Cuando interactuamos con otras personas o instituciones, desplegamos nuestro

conocimiento social y cultural sobre las variantes del castellano, y lo ponemos en juego de forma bastante consciente.

Cuando tratamos de adaptar nuestra forma de hablar para que una máquina reconozca lo que queremos decir, nuestra postura puede ser diferente de cuando interactuamos con personas. Muchas veces descartamos cuestionamientos que quizás sí plantearíamos a una persona o institución al encontrarnos en un contexto mediado por una tecnología como la inteligencia artificial, compleja y prestigiosa, pero también con limitaciones. Nos damos cuenta de que el sistema solo funciona bien si hablamos de cierta forma, pero no le atribuimos una intencionalidad, sino que asumimos que se trata de un mecanismo objetivo y simplemente tratamos de adaptarnos a sus limitaciones como algo no intencional. Efectivamente, hasta donde sabemos, las máquinas no tienen voluntad propia, pero el contexto de mediación tecnológica, tan nuevo, tan complejo, tan rodeado de grandes prestigios y grandes expectativas, dificulta que entendamos qué voluntades pueden estar involucradas en esa tecnología, más allá de la máquina que la implementa.

Pero incluso si no llegamos a identificar las voluntades involucradas en el desarrollo de las tecnologías que encontramos en nuestras vidas, sí podemos observar y entender el efecto de estas tecnologías. Por ejemplo, ¿qué efectos puede tener que adaptemos nuestra forma de hablar para facilitar que una máquina reconozca nuestras palabras? Puede suceder, que empecemos a considerar que nuestra variante del castellano no es moderna, no está alineada con el progreso tecnológico, no nos sirve para tener éxito en el mundo actual. Puede suceder que eso nos lleve a relegar nuestra variante materna, con la consiguiente pérdida de capacidad expresiva e incluso de identidad. Puede ser, también, que si no conseguimos adaptar nuestro dialecto, la máquina no reconozca lo que queremos decir, y eso puede tener efectos todavía más profundos: podemos sentirnos inútiles, incapaces de funcionar con éxito en el mundo actual. Puede contribuir a una imagen de nosotros mismos como ineptos que termine convirtiéndose en un obstáculo para

proyectarnos y funcionar de forma satisfactoria en una sociedad cada vez más mediada por tecnología.

Entonces, el comportamiento de un sistema de inteligencia artificial puede tener efectos mucho más allá de la simple interacción puntual entre la persona y la máquina. Si bien es cierto que resulta difícil predecir todos los efectos que puede tener una determinada tecnología en algo tan complejo como su uso en una sociedad, también es cierto que esa responsabilidad recae especialmente sobre los equipos que conciben, desarrollan e implementan esas tecnologías, ya que son los que las conocen mejor. Profundicemos un poco en cómo podemos empezar a abordar estas complejidades.

## SI ES SISTEMÁTICO NO ES ERROR, ES SESGO

Hemos dicho que no le atribuimos intencionalidad a la máquina, y todo parece indicar que, efectivamente, las máquinas no tienen intenciones. Pero la concepción, desarrollo y despliegue de la máquina están determinados por intenciones de grupos humanos.

En varias ocasiones hemos visto cómo los responsables de algunos sistemas de inteligencia artificial piden disculpas por efectos perjudiciales imprevistos de los sistemas que desarrollan. Por ejemplo, en el documental «El dilema de las redes sociales»[8], algunos de los entrevistados, personas involucradas en el desarrollo de estas tecnologías, explican que nunca imaginaron los efectos perniciosos que resultaron teniendo las redes sociales en cuanto a adicciones, su impacto en salud mental (por ejemplo, aumentando el índice de suicidios entre adolescentes), entre muchos otros aspectos.

En las disculpas, estos efectos perjudiciales se presentan como errores no intencionales. Sin embargo, en otro documental, «Prejuicio cifrado»[9], se describe cómo los efectos de muchos sistemas

8 «El dilema de las redes sociales» es un documental combinado con elementos de ficción, estrenado en 2020, dirigido por Jeff Orlowski y escrito por Orlowski, Davis Coombe, y Vickie Curtis.

9 «Prejuicio cifrado» es un documental dirigido por Shalini Kantayya y estrenado en 2020.

que involucran inteligencia artificial, especialmente los basados en datos (es decir, los que utilizan algoritmos de aprendizaje automático como los que discutimos en capítulos anteriores), son el producto de los prejuicios de sus creadores. En este documental se muestra cómo un sistema de reconocimiento de imágenes identifica con gran precisión rostros de personas con piel clara pero comete muchos más errores si se encuentra ante el rostro de una persona de piel más oscura.

Vemos aquí una gran diferencia entre un error accidental y un error sistemático. En el caso de un error sistemático, incluso si no es intencional o ni siquiera consciente, los efectos son también sistemáticos y, por lo tanto, pueden ser identificados, solucionados y, en el mejor de los casos, también prevenidos.

En los últimos tiempos hemos observado sistematicidades preocupantes en los errores de algunos sistemas de inteligencia artificial. Hemos entendido que los errores afectan de forma más perniciosa a personas de grupos minorizados, mientras que tienen un mejor funcionamiento con respecto a las personas de grupos dominantes. Por ejemplo, un sistema de filtrado automático de candidatos para lugares de trabajo para Amazon descartaba sistemáticamente a mujeres.[10] Twitter creaba automáticamente recortes de imágenes grandes en las que sistemáticamente se mostraban las caras de las personas en la imagen, priorizando personas blancas por encima de personas de pieles más oscuras.[11] El servicio de traducción automática de Google traduce los nombres de profesiones que tienen género neutro en inglés al género estereotipado para esas profesiones en castellano, contribuyendo de esta forma a reforzar estereotipos de género y a dificultar el acceso a determinadas profesiones para grandes sectores de la población.[12] Por ejemplo, se traducen *doctor* y

10 Más información sobre el caso de Amazon en: https://www.bbc.com/mundo/noticias-45823470.

11 Más información sobre el caso de Twitter en: https://www.vialibre.org.ar/twitter-investigara-si-su-algoritmo-tiene-un-sesgo-racial/

12 Más información sobre el caso de Google en: https://computerhoy.com/noticias/tecnologia/sesgo-genero-traductor-google-persiste-ella-cose-conduce-834637.

*nurse*, palabras que pueden aplicarse a personas de cualquier género en inglés, sistemáticamente como *doctor* y *enfermera* en castellano.

Este tipo de comportamientos es especialmente grave si tenemos en cuenta que estos sistemas tienen injerencia en derechos humanos fundamentales como educación, salud o justicia. Por ejemplo, un sistema de estimación de calificaciones escolares de Reino Unido asignó notas más bajas de lo que realmente habrían obtenido a personas de barrios de renta baja, pero hizo estimaciones más acordes con el resultado final para personas de barrios con mayor renta per cápita. Varios cuerpos de policía alrededor del mundo usan sistemas de reconocimiento facial para vigilar los espacios públicos y encontrar personas con orden de búsqueda y captura, pero como hemos mencionado más arriba, estos sistemas poseen menos exactitud al clasificar personas de pieles más oscuras que en personas de pieles más claras. En la justicia del estado de Florida, en Estados Unidos, un sistema que determinaba el riesgo de reincidencia en personas que solicitan libertad condicional estimaba un riesgo mayor al real para personas tipificadas como de etnia negra, y un riesgo menor al real para personas tipificadas como de etnia blanca.

A este tipo de comportamiento sistemático se lo conoce como sesgo, porque proviene de la intervención humana en la creación del sistema. Incluso en sistemas que no han sido programados explícitamente por personas, como los basados en aprendizaje automático, el sesgo humano determina con qué datos se entrena un modelo, cómo se representan esos datos, e incluso con qué algoritmo se infiere el modelo. Todas esas decisiones humanas, y por lo tanto subjetivas y con el sesgo propio de cada persona, contribuyen a la configuración del modelo final, y por lo tanto a su comportamiento. Por ejemplo, los grandes modelos de lenguaje que subyacen a los sistemas de traducción automática reproducen estereotipos de género y etnia, pero no reproducen lenguaje sexual explícito. Los datos con los que han sido entrenados no representan indistintamente todas las producciones lingüísticas, sino que muchas de ellas han sido vetadas según los valores de los equipos que los han creado y los grupos sociales que los financian.

## PODEMOS TRATAR EL SESGO

Por su sistematicidad, estos sesgos se pueden detectar con métricas bien establecidas, las llamadas *métricas de equidad* (*fairness* en inglés), siempre que se haya identificado el grupo social al que se está discriminando. Este grupo social se representa mediante uno o más atributos protegidos. Mediante estos atributos, las métricas de equidad describen con precisión si las predicciones de un modelo se distribuyen de forma indistinta entre la población que tiene el atributo protegido y la que no lo tiene. De esta forma se puede detectar si un sistema está actuando de forma discriminatoria con respecto a un grupo social que ya hemos identificado como vulnerable. Sin embargo, resulta mucho más complejo identificar comportamientos dañinos si no hemos identificado previamente a quiénes pueden afectar de forma sistemática. En el ejemplo con el que iniciábamos este capítulo no resulta fácil caracterizar las personas que pueden verse afectadas porque el sistema no reconoce sus palabras: puede tratarse de personas de ciertas regiones, pero también de ciertos grupos sociales, con voces más agudas o más graves, con ciertas particularidades neurológicas o motoras.

Afortunadamente, las métricas de equidad no son la única forma de inspeccionar el comportamiento de un sistema de inteligencia artificial. En los sistemas programados explícitamente se puede revisar el código para obtener una descripción de las acciones que podría llevar a cabo el sistema. Pero los sistemas basados en aprendizaje automático suelen producir modelos que resultan muy difíciles de comprender para los seres humanos. Sin embargo, se pueden aplicar mecanismos para que esos modelos ofrezcan, además de una predicción, también una explicación de las razones en las que se basa esa predicción. En esas explicaciones se pueden detectar razones inaceptables para nuestra sociedad, como por ejemplo la discriminación por etnia o género.

Dada la gravedad de estos efectos dañinos, sería muy importante poder prevenirlos en el momento de diseñar un sistema, en lugar

de detectarlos recién cuando el sistema ya está funcionando en la sociedad y afectando la vida de las personas. La principal dificultad para prevenirlos está en nuestras propias limitaciones cognitivas. El sesgo es un mecanismo cognitivo básico de los seres humanos, y resulta invisible para las personas que lo tienen. Por lo tanto, es prácticamente inevitable que un sistema diseñado por personas incorpore el sesgo de esas mismas personas. ¿Cómo hacer, entonces, para evitar los comportamientos dañinos sistemáticos? La mejor propuesta que tenemos hasta el momento no consiste en evitar los sesgos, sino en multiplicarlos. Podemos multiplicar las miradas diferentes que intervienen en el diseño de un sistema, o incluso antes, en el planteo de un problema, de una solución, de un producto o un servicio. Diferentes miradas pueden identificar problemas que resultan invisibles para quienes comparten un mismo sesgo, y, de esta forma, podemos intentar construir sistemas más respetuosos con las diversidades.

## ATENCIÓN: ¡INTELIGENCIA ARTIFICIAL EN CONSTRUCCIÓN!

En este capítulo hemos visto cómo los sistemas de inteligencia artificial incorporan los sesgos propios de sus creadores, y por esta razón pueden llegar a tener errores sistemáticos con efectos discriminatorios.

Ante la sistematicidad de los errores, los responsables de estos sistemas muchas veces alegan que los modelos predictivos sencillamente están reproduciendo las tendencias estadísticas que encontraron en los datos con los que fueron entrenados. Es decir, que los sesgos de los sistemas no se originan en los equipos que los crearon, sino que son tendencias propias de la sociedad. Sin embargo, al inspeccionar estos comportamientos en detalle, observamos que las sistematicidades encontradas se alinean con los valores de los grupos sociales dominantes que idean y financian estas tecnologías a más alto nivel, y no necesariamente con los fenómenos que efectivamente ocurren en la sociedad.

En cualquier caso, si el comportamiento de los sistemas es pernicioso, independientemente de cuáles sean las razones por las

que llegó a serlo, es necesario remediarlo. Contamos con leyes que garantizan muchos derechos fundamentales, como el derecho a la no discriminación, pero resulta difícil aplicar estos principios generales a casos concretos, y a menudo sutiles, que involucran tecnologías sofisticadas en interacciones complejas con la sociedad. Afortunadamente, en muchos países y también a nivel internacional, se están diseñando normativas específicas que determinan las responsabilidades, proveen mecanismos de control o disponibilizan canales para recibir las quejas y comentarios de los usuarios, de la misma forma que se desarrolló para otras áreas como alimentos, productos farmacéuticos, o derecho del consumidor en general. Resultan especialmente esperanzadoras las regulaciones que exigen la auditabilidad de los sistemas que impactan en derechos fundamentales de las personas, como la ley *rider* en España[13].

Estas exigencias regulatorias resultan totalmente factibles a nivel técnico. Afortunadamente, el área Inteligencia Artificial Responsable se ha desarrollado mucho en los últimos años y hoy contamos con herramientas como métricas para supervisar el comportamiento de los sistemas, metodologías para obtener explicaciones sobre las predicciones de los modelos y entornos de trabajo que facilitan estas herramientas.

Queremos cerrar este capítulo con un llamado a la acción. A pesar de las complejidades técnicas, los conceptos fundamentales en los que se basan los modelos de inteligencia artificial son intuitivos. También podemos comprender sin mucha dificultad cómo se comportan estos sistemas, aunque desconozcamos el detalle de cómo funcionan internamente, y tenemos instrumentos para detectar los sesgos. De esta forma, podemos convertirnos en agentes de cambio, ser una parte activa para mejorar estos sistemas y ayudar a construir una inteligencia artificial mejor para todos.

13 La ley *rider* (del inglés, *ciclista*) establece una serie de medidas de protección a los derechos laborales de las personas que se dedican al reparto domiciliario a través de plataformas digitales en España (2021).

# ÍNDICE

COLECCIÓN **KUAA**
dirigida por Federico Ariel

Es la colección de divulgación de la ciencia de Vera Cartonera. Con el vocablo en guaraní que se refiere al saber y al conocer, esta colección propone reunir a los protagonistas del quehacer científico para compartir, en palabras simples, cómo desde el sur empujamos las fronteras del conocimiento.




**VERA** editorial cartonera
Centro de Investigaciones Teórico–Literarias de la Facultad de Humanidades y Ciencias de la Universidad Nacional del Litoral.
Instituto de Humanidades y Ciencias Sociales IHUCSO Litoral (UNL/Conicet).
Programa de Lectura Ediciones UNL.












*Directora Vera cartonera:* Analía Gerbaudo

*Asesoramiento editorial:* Ivana Tosti

*Corrección editorial:* Laura Kiener y Valentina Miglioli

*Diseño:* Julián Balangero

Este libro fue compuesto con los tipos Alegreya y Alegreya Sans, de Juan Pablo del Peral *(www.huertatipografica.com)*.

---

¿Aprendizaje automágico? : Un viaje al corazón de la inteligencia artificial contemporánea / Enzo Ferrante ... [et al.] ; dirigido por Enzo Ferrante ; prólogo de Diego Golombek. - 1a ed. - Santa Fe : Universidad Nacional del Litoral, 2022.
Libro digital, PDF/A - (Vera Cartonera / Analía Gerbaudo ; Kuaa)
Archivo Digital: descarga y online

ISBN 978-987-692-317-0

1. Inteligencia Artificial. 2. Tecnología Informática. 3. Ciencias Tecnológicas. I. Ferrante, Enzo, dir. II. Golombek, Diego, prolog.
CDD 006.301

---





Facultad de Humanidades y Ciencias UNL
Ciudad Universitaria, Santa Fe, Argentina
Contacto: veracartonera@fhuc.unl.edu.ar